Теоретико-групповой подход в комбинаторной теории обобщающей способности

Фрей Александр Ильич

Московский физико-технический институт
(Государственный университет)
Кафедра «Интеллектуальные Системы» (ВЦ РАН)

Научный руководитель: к.ф.-м.н. Воронцов Константин Вячеславович

21 апреля 2011

Общая постановка задачи

- Исследуется функционал полного скользящего контроля
  - Предполагается стационарность данных
  - Фиксируется некоторое семейство алгоритмов (например, линейные разделяющие поверхности)
  - Метод обучения – минимизация эмпирического риска
- Цели:
  - Получение новых методов регуляризации (контроля сложности семейства), учитывающих расслоение и связность

Карта линейных разделяющих правил

- Выборка $X_L \subset \mathbb{R}^3$
- Правила вида $y = \mathrm{sign}(< w, x >)$
- Точки сферы — направляющие векторы линейной разделяющей поверхности
- Грани графа на сфере — классы эквивалентных алгоритмов
- ToDo: вставить картинку.

Профиль r-связности

- Задача бинарной классификации: $Y = \{1, -1\}$
- $S_2 = \{e, h\}: Y \to Y$
- Группа $G = (S_2)^L$, элемент $g \in G = \{h_1, h_2, \dots, h_L\}$

$$(ga)(x_i) = \begin{cases} a(x_i), & \text{при } h_i = e \\ 1 - a(x_i), & \text{при } h_i \neq e. \end{cases}$$

- Действие G — изометрия (по расстоянию Хэмминга)
- Профиль r-связности A — функцию от параметра q:

$$\Theta_r(q, A) = \sum_{a \in A} [|B_r(a, A)| = q].$$

Значение $\Theta_r(q, A)$ соответствует числу алгоритмов $a \in A$, имеющих ровно q соседей в окрестности $B_r(a, A)$.

Профиль расслоения-связности

- Профиль расслоения $\Delta(m, A)$:

$$\Delta(m, A) = \sum_{a \in A} [n(a, \mathbb{X}) = m].$$

- Профиль расслоения-связности $\Lambda_r(m, q, A)$ семейства A:

$$\Lambda_r(m, q, A) = \Bigl| \{a \in A \colon |B_r(a, A)| = q \text{ и } n(a, \mathbb{X}) = m\} \Bigr|.$$

(число алгоритмов $a \in A$ с m ошибками и q соседями в окрестности радиуса r)

- Теорема о наиболее вероятном профиле расслоения:

$$\frac{1}{|G|} \sum_{g \in G} \Delta(m, gA) = \frac{C_L^m}{2^L} |A|.$$

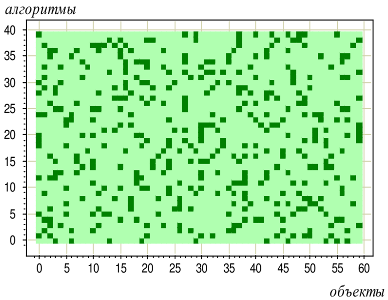
- Теорема о профиле расслоения-связности

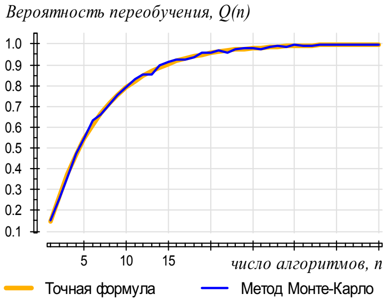
$$\frac{1}{|G|} \sum_{g \in G} \Lambda_r(m, q, gA) = \frac{C_L^m}{2^L} \cdot \Theta_r(q, A)$$

Результаты, полученные для РМЭР

- **Связка монотонных цепочек**
  - Фрей А. И., Точные оценки вероятности переобучения для симметричных семейств алгоритмов // Всеросс. конф. ММРО-14 — М.: МАКС Пресс, 2009. — С. 66–69.
- **Шар алгоритмов и центральный слой шара**
  - Толстихин И. О., Точная оценка вероятности переобучения для одного специального семейства алгоритмов // Конференция «Ломоносов-2010».
- **Полный слой и полный куб алгоритмов**
  - Frei A. I., Accurate Estimates of the Generalization Ability for Symmetric Sets of Predictors and Randomized Learning Algorithms // Pattern Recognition and Image Analysis. — 2010. — Vol. 20, No. 3. — Pp. 241-250.
- **Монотонные и унимодальные сетки**
  - Фрей А. И., Вероятность переобучения плотных и разреженных семейств многомерных сеток алгоритмов // Международ. конф. ИОИ-8 — М.: МАКС Пресс, 2010. — С. 87–90.

Отсутствие связности: случайные множества алгоритмов





- Пусть A_m^n — множество из n алгоритмов, каждый из которых допускает m ошибок на полной выборке. Векторы ошибок независимы.

Теорема (Вероятность переобучения для A_m^n)

Пусть μ — рандомизированный МЭР. Тогда

$$\mathbf{E}_G Q_\mu(\varepsilon, A_m^n) = 1 - (1 - Q_\mu(\varepsilon, a_m))^n$$

Отсутствие связности: случайные множества алгоритмов

- Пронумеруем алгоритмы: $A = (a_1, \dots, a_d)$
- S_L — симметрическая группа порядка L;
  - S_L действует на объектах выборки;
  - S_L действует на векторах ошибок алгоритмов;
- $G = (S_L)^d$ — свободное произведение S_L;
  - $g = (g_1, \dots, g_d) \in G$ — элемент группы G, $g_i \in S_L$;
  - G действует на векторе алгоритмов:

$$gA = \bigl(g_1(a_1), \dots, g_d(a_d)\bigr).$$

- Вероятность переобучения несвязной перестановки A:

$$\bar{Q}_\mu(\varepsilon, A) = \mathbf{E}_G\, Q_\mu(\varepsilon, gA), \;\text{где}\; \mathbf{E}_G \stackrel{def}{=} \frac{1}{(L!)^d} \sum_{g \in G}$$

- $\bar{Q}_\mu(\varepsilon, A)$ зависит только от профиля расслоения A.